Weak Identification in Fuzzy Regression Discontinuity Design*

Thomas Lemieux[†] Vadim Marmer[†]

May 26, 2010

Abstract

We consider weak identification in the fuzzy regression discontinuity (FRD) model. In this model, the treatment effect is identified through a discontinuity in the conditional probability of treatment assignment. Weak identification corresponds to the situation where the discontinuity is of a small magnitude. When identification is weak, we show that the usual t-test based on the FRD estimator and its standard error suffers from asymptotic size distortions. To eliminate those size distortions, we propose a modified t-statistic that uses a null-restricted version of the standard error of the FRD estimator. Simple and asymptotically valid confidence sets for the treatment effect can be also constructed using the FRD estimator and its the null-restricted standard error.

JEL Classification: C12; C13; C14
Keywords: Nonparametric inference; regression discontinuity design; treatment effect; weak identification

1 Introduction

In this paper, we discuss the problem of weak identification in the context of fuzzy regression discontinuity (FRD) design. The regression discontinuity (RD) design has

*We thank Taisuke Otsu and Eric Renault for helpful comments.
[†]Department of Economics, University of British Columbia, 997 - 1873 East Mall, Vancouver, BC V6T 1Z1, Canada. E-mail: tlemieux@interchange.ubc.ca (Lemieux) and vadim.marmer@ubc.ca (Marmer).

been studied recently by Hahn, Todd, and Van der Klaauw (2001), HTV hereafter, and Imbens and Lemieux (2008), IL hereafter. The RD framework is concerned with evaluating the effects of interventions or treatments when assignment to treatment is determined completely or partly by the value of an observable predictor variable. In this framework, identification of the treatment effect comes from the discontinuity in the conditional probability function of treatment assignment given the value of the predictor, where the discontinuity occurs at some known point. When assignment is completely determined by the value of the predictor, the RD design is called sharp. The focus of this paper is the fuzzy RD design which occurs when the predictor determines assignment only partly.

Weak identification in FRD corresponds to the situation where the discontinuity in the conditional probability function of treatment assignment is of a small magnitude. Similarly to the weak instruments literature (see, for example, Andrews and Stock (2007) for a review), weak identification can be formally modelled using the local-to-zero framework. Specifically, we assume that the discontinuity in the conditional probability function of treatment assignment is local-to-zero.

When identification is weak, we show that the usual t-test based on the FRD estimator and its standard error suffers from asymptotic size distortions with an exception to a few specific situations. For example, one can still use the usual t-statistic when testing the hypothesis of zero treatment effect if the assignment to treatment and outcome variables are asymptotically independent. However, in general the usual t-test is asymptotically invalid because it can overreject the null hypothesis when identification is weak. The usual confidence intervals constructed as estimate $\pm$ constant $\times$ standard error are also invalid because their asymptotic coverage probability can be below the assumed nominal coverage when identification is weak.

In this paper, we suggest a simple modification to the t-test that eliminates the asymptotic size distortions caused by weak identification. Unlike the usual t-statistic, the proposed modified t-statistic uses a null-restricted version of the standard error of the FRD estimator. Tests based on the t-statistic computed using the null-restricted standard errors do not suffer from asymptotic size distortions when identification is weak and are asymptotically equivalent to the usual t-test when identification is strong.

Asymptotically valid confidence sets for the treatment effect can be obtained by inverting the test based on the t-statistic with the null-restricted standard error. These confidence sets are easy to compute as their construction only involve solving a simple quadratic equation. Unlike the usual confidence intervals constructed as estimate $\pm$ constant $\times$ standard error, the confidence sets we propose can be unbounded with positive probability. This property, however, is expected from valid confidence sets in the situations with local identification failure and unbounded parameter space as is demonstrated in Dufour (1997).

In a recent paper, Otsu and Xu (2010), propose empirical likelihood based confidence sets for the RD. Their method does not involve variance estimation and for that reason is expected to be robust to weak identification. It however requires computation of the empirical likelihood function numerically and is computationally more demanding than our approach. At the same time, the empirical likelihood based confidence sets are expected to have better higher-order coverage properties.

We proceed as follows. In Section 2 we describe the FRD model and present our analytical results. In Section 3, we illustrate our results in a Monte Carlo experiment.

2 Theoretical results

2.1 Preliminaries

In the RD model, the observed outcome variable y_i is written as

$$y_i = y_{0i} + x_i \beta_i,$$

where x_i is the treatment indicator variable that takes on value one if treatment is received and x_i is equal to zero otherwise, y_{0i} is the outcome without treatment, and β_i is the random treatment effect for observation i. The treatment assignment depends on another observable variable, z_i:

$$\Pr\left(x_i = 1 | z_i = z\right) = E\left(x_i | z_i = z\right).$$

The main feature in this framework is that $E\left(x_i | z_i = z\right)$ is discontinuous at some known point z_0, while $E\left(y_{0i} | z_i\right)$ is assumed to be continuous at z_0.

Assumption 1. **(a)** $\lim_{z \downarrow z_0} E\left(x_i | z_i = z\right) \neq \lim_{z \uparrow z_0} E\left(x_i | z_i = z\right)$.

(b) $\lim_{z \downarrow z_0} E\left(y_{0i} | z_i = z\right) = \lim_{z \uparrow z_0} E\left(y_{0i} | z_i = z\right)$.

The RD design is called sharp if $\left|\lim_{z\uparrow z_0} E(x_i|z_i = z) - \lim_{z\downarrow z_0} E(x_i|z_i = z)\right| = 1$. In this case, the treatment assignment is completely determined by the value of z_i. The FRD design corresponds to the situation where

$$\left|\lim_{z\uparrow z_0} E(x_i|z_i = z) - \lim_{z\downarrow z_0} E(x_i|z_i = z)\right| < 1,$$

so either $0 < \lim_{z\uparrow z_0} E(x_i|z_i = z) < 1$ or $0 < \lim_{z\downarrow z_0} E(x_i|z_i = z) < 1$ or both, and therefore the treatment assignment is not a deterministic function of z_i.

The main object of interest is the RD effect

$$\beta = \frac{y^+ - y^-}{x^+ - x^-}, \tag{1}$$

where

$$\begin{aligned} y^+ &= \lim_{z\downarrow z_0} E(y_i|z_i = z), & x^+ &= \lim_{z\downarrow z_0} E(x_i|z_i = z), \\ y^- &= \lim_{z\uparrow z_0} E(y_i|z_i = z), & x^- &= \lim_{z\uparrow z_0} E(x_i|z_i = z). \end{aligned}$$

The exact interpretation of β depends on the assumptions that the econometrician is willing to make in addition to Assumption 1. As discussed in HTV, if β_i and x_i are assumed to be independent conditional on z_i, then β captures the ATE at $z_i = z_0$: $\beta = E(\beta_i|z_i = z_0)$. This also covers a special case where the treatment effect is a deterministic function of z_i in the neighborhood of z_0: $\beta_i = \beta(z_i)$. In this case, $\beta = \beta(z_0)$ and it is referred to in HTV as a constant treatment effect.

HTV show that another interpretation for β can be obtained if one assumes that in the neighborhood of z_0 and with probability one, x_i is a non-decreasing or non-increasing function of z_i, and $E(x_i\beta_i|z_i = z)$ is constant in the neighborhood of z_0. In this case, β captures the local ATE or the ATE for compliers, where compliers are observations i for which x_i switches its value from zero to one when z_i changes from $z_0 - e$ to $z_0 + e$ for some small $e > 0$.

Regardless of its interpretation, it is now standard to estimate β using the local linear approach. Define

$$\left(\hat{y}^+, \hat{b}_y^+\right) = \arg\min_{a,b} \sum_{i=1}^{n} \left(y_i - a - (z_i - z_0)\, b\right)^2 I_i^+ K\left(\frac{z_i - z_0}{h}\right),$$

$$
\begin{aligned}
\left(\hat{y}^-, \hat{b}_y^-\right) &= \arg\min_{a,b} \sum_{i=1}^n \left(y_i - a - (z_i - z_0)\, b\right)^2 I_i^- K\left(\frac{z_i - z_0}{h}\right), \\
\left(\hat{x}^+, \hat{b}_x^+\right) &= \arg\min_{a,b} \sum_{i=1}^n \left(x_i - a - (z_i - z_0)\, b\right)^2 I_i^+ K\left(\frac{z_i - z_0}{h}\right), \\
\left(\hat{x}^-, \hat{b}_x^-\right) &= \arg\min_{a,b} \sum_{i=1}^n \left(x_i - a - (z_i - z_0)\, b\right)^2 I_i^- K\left(\frac{z_i - z_0}{h}\right),
\end{aligned}
$$

where K is the kernel function, h is the bandwidth, and the indicator functions I_i^- and I_i^+ are defined as

$$
\begin{aligned}
I_i^- &= 1\left\{z_i \leq z_0\right\}, \\
I_i^+ &= 1 - 1\left\{z_i \leq z_0\right\}.
\end{aligned}
$$

The local linear estimator of β is given by

$$
\hat{\beta} = \frac{\hat{y}^+ - \hat{y}^-}{\hat{x}^+ - \hat{x}^-}.
$$

To describe its asymptotic behavior, consider the following high-level assumption.

Assumption 2. **(a)** *The PDF of z_i is continuous and bounded in the neighborhood of z_0; it is also bounded away from zero in the neighborhood of z_0.*

(b) *The data $\{(y_i, x_i, z_i)\}_{i=1}^n$, kernel function K, and bandwidth h are such that*

$$
\sqrt{nh}\begin{pmatrix} \hat{y}^+ - y^+ \\ \hat{x}^+ - x^+ \\ \hat{y}^- - y^- \\ \hat{x}^- - x^- \end{pmatrix} \rightarrow_d \begin{pmatrix} Y^+ \\ X^+ \\ Y^- \\ X^- \end{pmatrix} =^d N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{yy}^+ & \sigma_{yx}^+ & 0 & 0 \\ \sigma_{yx}^+ & \sigma_{xx}^+ & 0 & 0 \\ 0 & 0 & \sigma_{yy}^- & \sigma_{yx}^- \\ 0 & 0 & \sigma_{yx}^- & \sigma_{xx}^- \end{pmatrix}\right).
$$

(c) *There exist $\hat{\sigma}_{fg}^s$, $s \in \{+, -\}$ and $f, g \in \{x, y\}$, such that $\hat{\sigma}_{fg}^s \rightarrow_p \sigma_{fg}^s$ for all s and f, g.*

Remark. As discussed in HTV (pages 207-208) and IL (page 630), Assumption 2 is satisfied when, for example, the data $\{(y_i, x_i, z_i)\}_{i=1}^n$ are iid, K is a continuous, symmetric around zero, non-negative, and compactly supported second-order kernel function, $h = cn^{-\delta}$ with $c > 0$ and $1/5 < \delta < 2/5$, and provided that some additional technical conditions hold.[1] In this case,

$$\begin{aligned}
\sigma_{yy}^{+} &= \frac{k^{+}}{f_z(z_0)} \lim_{z \downarrow z_0} Var(y_i|z_i = z_0), \ \sigma_{yy}^{-} = \frac{k^{-}}{f_z(z_0)} \lim_{z \uparrow z_0} Var(y_i|z_i = z_0), \\
\sigma_{xx}^{+} &= \frac{k^{+}}{f_z(z_0)} \lim_{z \downarrow z_0} Var(x_i|z_i = z_0), \ \sigma_{xx}^{-} = \frac{k^{-}}{f_z(z_0)} \lim_{z \uparrow z_0} Var(x_i|z_i = z_0), \\
\sigma_{yx}^{+} &= \frac{k^{+}}{f_z(z_0)} \lim_{z \downarrow z_0} Cov(y_i, x_i|z_i = z_0), \ \sigma_{yx}^{-} = \frac{k^{-}}{f_z(z_0)} \lim_{z \uparrow z_0} Cov(y_i|z_i = z_0),
\end{aligned}$$

where $f_z(z_0)$ denotes the PDF of z_i at $z = z_0$,

$$k^{+} = \frac{\int_0^\infty \left(\int_0^\infty s^2 K(s)\, ds - u \int_0^\infty sK(s)\, ds\right)^2 K^2(u)\, du}{c \cdot \left(\int_0^\infty u^2 K(u)\, du \int_0^\infty K(u)\, du - \left(\int_0^\infty uK(u)\, du\right)^2\right)^2},$$

and k^{-} is defined similarly to k^{+} but with the integrals over $(-\infty, 0)$, see Theorem 4 in HTV.

The asymptotic variance σ_{yy}^{+} can be consistently estimated by

$$\hat{\sigma}_{yy}^{+} = \frac{k}{\hat{f}_z^2(z_0)} \frac{1}{nh} \sum_{i=1}^{n} \left(y_i - \hat{y}^{+}\right)^2 I_i^{+} K\left(\frac{z_i - z_0}{h}\right),$$

where $\hat{f}_z(z_0)$ is the kernel estimator of $f_z(z_0)$: $\hat{f}_z(z_0) = (nh)^{-1} \sum_{i=1}^{n} K((z_i - z_0)/h)$. Consistent estimators of $\sigma_{xx}^{+}, \sigma_{yx}^{+}, \sigma_{yy}^{-}, \sigma_{xx}^{-}, \sigma_{yx}^{-}$ can be constructed similarly.

When Assumption 2 holds and $p\lim_{n\to\infty}(\hat{x}^{+} - \hat{x}^{-}) \neq 0$, by a standard application of the delta-method, the asymptotic distribution of the FRD estimator of β is given by

$$\sqrt{nh}\left(\hat{\beta} - \beta\right) \to_d N(0, V_\beta), \tag{2}$$

where

$$V_\beta = \frac{1}{(x^{+} - x^{-})^2} \left(\sigma_{yy}^{+} + \sigma_{yy}^{-} + \beta^2 \left(\sigma_{xx}^{+} + \sigma_{xx}^{-}\right) - 2\beta \left(\sigma_{yx}^{+} + \sigma_{yx}^{-}\right)\right). \tag{3}$$

[1] The requirement $\delta > 1/5$ corresponds to data undersmoothing and is needed to eliminate the asymptotic bias of the kernel estimators.

The asymptotic variance V_β can be consistently estimated by the plug-in method with

$$\hat{V}_\beta = \frac{1}{\left(\hat{x}^+ - \hat{x}^-\right)^2}\left(\hat{\sigma}^+_{yy} + \hat{\sigma}^-_{yy} + \hat{\beta}^2\left(\hat{\sigma}^+_{xx} + \hat{\sigma}^-_{xx}\right) - 2\hat{\beta}\left(\hat{\sigma}^+_{yx} + \hat{\sigma}^-_{yx}\right)\right).$$

A test of $H_0 : \beta = \beta_0$ in practice is usually based on the t-statistic

$$T\left(\beta_0\right) = \frac{\hat{\beta} - \beta_0}{\sqrt{\hat{V}_\beta / (nh)}},$$

and one rejects H_0 when T exceeds standard normal critical values.

2.2 Weak identification in FRD

The FRD effect β is not defined if Assumption 1(a) fails and $x^+ - x^- = 0$. Here we consider the situation where β is well-defined however only weakly identified. The issue of weak identification arises in the FRD model when the conditional probability function of receiving the treatment has discontinuity only of a small magnitude. Similarly to the case of weak instruments in the IV regression model, this creates the problem of nearly zero denominator in (1) and (3). The consequence of the weak identification is that the asymptotic result in (2) provides a poor approximation to the actual behavior of the estimator in finite samples.

A useful device for analyzing the properties of estimators in the case of weak identification is local-to-zero asymptotics. We make the following assumption.

Assumption 3. *(**Weak ID**) We assume that the bandwidth h in Assumption 2 is such that $x^+ - x^- = \pi_n = \theta/\sqrt{nh}$ for some constant θ.*

Remarks. (a) The weak ID condition assumes that the function $E\left(x_i | z_i = z\right)$ has a discontinuity at z_0 of only a small magnitude. In the case of weak instruments, one usually assumes that in the first-stage equation, the coefficient of the IVs are local-to-zero: $\theta/\sqrt{n}$. Such an assumption results in a non-trivial asymptotic distribution for the IV estimator. In our case, due to nonparametric rates of convergence of the estimators, one has to consider the sequence π_n that converges to zero at $1/\sqrt{nh}$ rate.

(b) Note that the bandwidth h in the Weak ID condition is the bandwidth chosen by the econometrician for the estimation of x^+, x^-, y^+, and y^-. Thus, formally the assumption Weak ID states that the model depends on the sample size and the choice of the bandwidth. Intuitively, the assumption implies that the econometrician cannot achieve identification by simply estimating x^+, x^-, y^+, and $y-$ using a different bandwidth, say h^*, such that $h^*/h \to \infty$, i.e. we assume that the weak identification problem persists regardless of the bandwidth choice.

Theorem 1. *Under Assumptions 2 and 3, the following results hold jointly:*

(a) $\hat{\beta} - \beta \to_d \xi_{\beta,\theta}$, *where*

$$\xi_{\beta,\theta} = \frac{Y^+ - Y^- - \beta\left(X^+ - X^-\right)}{\left(X^+ - X^-\right) + \theta}.$$

(b) $(nh)^{-1}\hat{V}_\beta \to_d \sigma^2\left(\beta + \xi_{\beta,\theta}\right)/\left(X^+ - X^- + \theta\right)^2$, *where for* $b \in \mathbb{R}$,

$$\sigma^2(b) = \sigma_{yy}^+ + \sigma_{yy}^- + b^2\left(\sigma_{xx}^+ + \sigma_{xx}^-\right) - 2b\left(\sigma_{yx}^+ + \sigma_{yx}^-\right).$$

(c) *Under* $H_0 : \beta = \beta_0$, $|T(\beta_0)| \to_d |\mathcal{Z}_\beta|\left(\sigma(\beta)/\sigma\left(\beta + \xi_{\beta,\theta}\right)\right)$, *where*

$$\mathcal{Z}_\beta = \frac{Y^+ - Y^- - \beta\left(X^+ - X^-\right)}{\sigma(\beta)} \sim N(0,1).$$

Remarks. (a) Part (a) of the theorem shows that, due to weak identification, the FRD estimator $\hat{\beta}$ is inconsistent.

(b) According to part (b) of the theorem, the estimator of the asymptotic variance of FRD $\hat{V}_\beta$ diverges at the rate nh due to the presence of $(\hat{x}^+ - \hat{x}^-)^2$ in the denominator. However the standard error $\sqrt{\hat{V}_\beta/(nh)}$ is stochastically bounded in large samples.

(c) The asymptotic distribution of the t-statistic in part (c) is nonstandard. The marginal distribution of $\mathcal{Z}_\beta$ is standard normal, however, the random variables $\mathcal{Z}_\beta$ and $\sigma\left(\beta + \xi_{\beta,\theta}\right)$ are not independent.

Consider a test of $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$ with the nominal asymptotic size α based on the usual t-statistic. The econometrician rejects H_0 when $|T(\beta_0)| > z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. The true asymptotic size is given by

$$\lim_{n\to\infty} P\left(|T(\beta_0)| > z_{1-\alpha/2} \mid \beta = \beta_0\right) = P\left(|\mathcal{Z}_\beta| \frac{\sigma(\beta)}{\sigma\left(\beta + \xi_{\beta,\theta}\right)} > z_{1-\alpha/2} \mid \beta = \beta_0\right).$$

If $\sigma(\beta)/\sigma\left(\beta+\xi_{\beta,\theta}\right)\leq 1$ with probability one, the true asymptotic size is less or equal to α . In such a case, the test based on $T(\beta_0)$ is asymptotically valid. For example, suppose that $\beta=0$ and selection into treatment and outcome are asymptotically independent, i.e. $\sigma_{yx}^{+}=\sigma_{yx}^{-}=0$. Since

$$\frac{\sigma^2(\beta)}{\sigma^2\left(\beta+\xi_{\beta,\theta}\right)}=\left(1+\left(\xi_{\beta,\theta}^2+2\beta\xi_{\beta,\theta}\right)\frac{\sigma_{xx}^{+}+\sigma_{xx}^{-}}{\sigma^2(\beta)}-2\xi_{\beta,\theta}\frac{\sigma_{yx}^{+}+\sigma_{yx}^{-}}{\sigma^2(\beta)}\right)^{-1},$$

we obtain that in this case, $\sigma(\beta)/\sigma\left(\beta+\xi_{\beta,\theta}\right)\leq 1$ with probability one, and consequently the test based on $T(\beta_0)$ is conservative: $\lim_{n\to\infty}P\Big(|T(\beta_0)|>z_{1-\alpha/2}\big|\beta=\beta_0\Big)\leq\alpha$.

If on the other hand $\sigma(\beta)/\sigma\left(\beta+\xi_{\beta,\theta}\right)>1$ with high probability, one can expect asymptotic size distortions, i.e. $\lim_{n\to\infty}P\Big(|T(\beta_0)|>z_{1-\alpha/2}\big|\beta=\beta_0\Big)>\alpha$, and that the usual confidence intervals constructed as $\hat{\beta}\pm z_{1-\alpha/2}\times\sqrt{\hat{V}_\beta/(nh)}$ will have the asymptotic coverage probability less than their nominal coverage $1-\alpha$. For example, asymptotic size distortions are more likely to occur when the selection into treatment variable x_i and the outcome variable y_i are highly correlated. Note that for the IV regression model, substantial size distortions are reported when the instruments are weak and the correlation between endogenous regressors and instruments is high (Staiger and Stock, 1997, page 577).

2.3 Weak identification robust inference for FRD

As it is apparent from Theorem 1, the failure of the standard t-test when identification is weak is due to asymptotic behavior of $\hat{V}_\beta$ which depends on inconsistent estimator $\hat{\beta}$. Inconsistency of $\hat{\beta}$ generates an extra term $\sigma(\beta)/\sigma\left(\beta+\xi_{\beta,\theta}\right)$ for the asymptotic distribution of the t-statistic T.

Instead of $\hat{V}_\beta$, we suggest using the following null-restricted estimator of the asymptotic variance. Consider $H_0:\beta=\beta_0$ and

$$\begin{aligned}\tilde{V}_\beta(\beta_0) &= \frac{\hat{\sigma}^2(\beta_0)}{\left(\hat{x}^{+}-\hat{x}^{-}\right)^2},\text{ where}\\ \hat{\sigma}^2(b) &= \hat{\sigma}_{yy}^{+}+\hat{\sigma}_{yy}^{-}+b^2\left(\hat{\sigma}_{xx}^{+}+\hat{\sigma}_{xx}^{-}\right)-2b\left(\hat{\sigma}_{yx}^{+}+\hat{\sigma}_{yx}^{-}\right).\end{aligned}$$

Further, consider a null-restricted version of the t-statistic based on $\tilde{V}(\beta_0)$:

$$\tilde{T}(\beta_0) = \frac{\hat{\beta} - \beta_0}{\sqrt{\tilde{V}_\beta(\beta_0)/(nh)}}.$$

Theorem 2. *Let $\mathcal{Z} \sim N(0,1)$. Under Assumptions 2 and 3, and for a fixed constant $\delta = \beta - \beta_0$, $\left|\tilde{T}(\beta_0)\right| \to_d \left|\mathcal{Z} + \frac{\theta\delta}{\sigma(\beta_0)}\right|$.*

Remarks. (a) The t-statistic with a null-restricted variance estimator has a standard normal asymptotic distribution under $H_0 : \beta = \beta_0$. For fixed alternatives $\beta = \beta_0 + \delta$, the asymptotic distribution of $\tilde{T}(\beta_0)$ is noncentral, and one can expect nontrivial power against such alternatives. As usual in the case of weak identification, there is no power against local alternatives $\beta = \beta_0 + \delta/\sqrt{nh}$, as in this case $\tilde{T}(\beta_0) \to_d \mathcal{Z}$ for all values of δ. Power of the test depends on the strength of identification θ and the distance from the null δ.

(b) It is easy to show that the result in part (b) remains unchanged in the case of strong identification $x^+ - x^- = \theta$ and local alternatives $\beta = \beta_0 + \delta/\sqrt{nh}$. Thus, when identification is strong, a test based on $\tilde{T}(\beta_0)$ has nontrivial power against local alternatives.

While the usual t-test can have size distortions when identification is weak, a test based on the t-statistic with the null-restricted standard error remains asymptotically valid whether identification is weak or strong. The test is to reject H_0 when $\left|\tilde{T}(\beta_0)\right| > z_{1-\alpha/2}$.

One can construct a confidence set for β with asymptotic coverage probability $1-\alpha$ by collecting the values β_0 that cannot be rejected by the $\tilde{T}(\beta_0)$ test:

$$CS_{1-\alpha} = \left\{\beta_0 \in R : \left|\tilde{T}(\beta_0)\right| \leq z_{1-\alpha/2}\right\}. \tag{4}$$

The confidence set $CS_{1-\alpha}$ can be easily computed analytically by solving for the values of β_0 that satisfy the inequality

$$nh(\hat{\beta} - \beta_0)^2(\hat{x}^+ - \hat{x}^-)^2 - z^2_{1-\alpha}(\hat{\sigma}_{yy} + \beta_0^2\hat{\sigma}_{xx} - 2\hat{\sigma}_{yx}\beta_0) \leq 0, \tag{5}$$

where $\hat{\sigma}_{yy} = \hat{\sigma}^+_{yy} + \hat{\sigma}^-_{yy}$, $\hat{\sigma}_{xx} = \hat{\sigma}^+_{xx} + \hat{\sigma}^-_{xx}$, and $\hat{\sigma}_{yx} = \hat{\sigma}^+_{yx} + \hat{\sigma}^-_{yx}$.

The expression on the left-hand side in (5) is a second order polynomial in β_0.

Depending on the coefficients of that polynomial, the confidence set $CS_{1-\alpha}$ potentially can take one of the following forms: (i) an interval, (ii) the entire real line, or (iii) a union of two disconnected half-lines $(-\infty, a_1] \cup [a_2, \infty)$, where $a_1 < a_2$.[2] It is equal to the entire real line when the discriminant and the coefficient on β_0^2 in the quadratic polynomial in β_0 in (5) are both negative. The discriminant of that polynomial is negative when $nh(\hat{x}^+ - \hat{x}^-)^2(\hat{\beta}^2\hat{\sigma}_{xx} - 2\hat{\beta}\hat{\sigma}_{yx} + \hat{\sigma}_{yy}) - z_{1-\alpha/2}^2\left(\hat{\sigma}_{xx}\hat{\sigma}_{yy} - \hat{\sigma}_{yx}^2\right) < 0$, and the coefficient on β_0^2 is negative when $nh(\hat{x}^+ - \hat{x}^-)^2 - z_{1-\alpha/2}^2\hat{\sigma}_{xx} < 0$. When identification is strong and as the sample size n increases, both the discriminant and the coefficient on β_0^2 tend to be positive, and therefore, with probability approaching one, $CS_{1-\alpha}$ will be an interval when identification is strong.

When identification is weak, however, $nh(\hat{x}^+ - \hat{x}^-)^2$ approaches a constant θ and $\hat{\beta} \to_d \beta + \xi_{\beta,\theta}$ as n increases. In this case, the confidence set $CS_{1-\alpha}$ can be unbounded with a positive probability. This probability depends on the strength of identification θ and is higher for smaller values of $|\theta|$. Thus, when $\theta = 0$, the confidence set $CS_{1-\alpha}$ is equal to the entire real line with probability approaching one.

2.4 Relation to the Anderson-Rubin test

The test based on $\tilde{T}(\beta_0)$ can be given an interpretation of the Anderson-Rubin (AR) test (Anderson and Rubin, 1949; Dufour, 1997; Staiger and Stock, 1997). The idea behind the AR test is to avoid estimation of β and to rely upon the null-restricted sample moment conditions. In the case of FRD, the identifying population condition for β is

$$y^+ - y^- - \beta\left(x^+ - x^-\right) = 0. \tag{6}$$

Consider again $H_0 : \beta = \beta_0$. Following the AR approach, one should consider the asymptotic distribution of the re-scaled sample counterpart of the expression in (6):

$$S(\beta_0) = \sqrt{nh}\left(\hat{y}^+ - \hat{y}^- - \beta_0\left(\hat{x}^+ - \hat{x}^-\right)\right).$$

Under H_0, (6) holds with $\beta = \beta_0$, and therefore $S(\beta_0)$ is correctly centered around zero (asymptotically):

$$S(\beta_0) = \sqrt{nh}\left(\hat{y}^+ - \hat{y}^- - \left(y^+ - y^-\right) - \beta_0\left(\hat{x}^+ - \hat{x}^- - \left(x^+ - x^-\right)\right)\right).$$

[2] We show in the appendix that the confidence set $CS_{1-\alpha}$ cannot be empty.

Since the asymptotic variance of $S(\beta_0)$ is given by $\sigma^2(\beta_0)$, and $\hat{\sigma}^2(\beta_0)$ is a consistent estimator of $\sigma^2(\beta_0)$, the resulting asymptotically pivotal AR statistic is then $S(\beta_0)/\hat{\sigma}(\beta_0)$, however, $S(\beta_0)/\hat{\sigma}(\beta_0) = \tilde{T}(\beta_0)$.

3 Monte Carlo experiment

In this section, we illustrate the problem of weak identification in the FRD model by a Monte Carlo experiment. In our experiment, the outcome variable y_i is generated according to the following model:

$$
\begin{aligned}
y_i &= y_{0i} + x_i\beta, \\
x_i &= \begin{cases} 1(u_i < 0), & z_i \leq 0, \\ 1(u_i < c), & z_i > 0, \end{cases}
\end{aligned}
$$

where the outcome without treatment variable y_{0i} and the variable u_i are bivariate normal:

$$
\begin{pmatrix} y_{0i} \\ u_i \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).
$$

Note that in this setup, u_i determines whether the treatment is received, and therefore the parameter ρ captures the degree of endogeneity of the treatment in this experiment. The treatment predictor variable z_i is generated to be standard normal and independent of y_{0i}, u_i. The observations are simulated to be independent across i's.

Let $\Phi(\cdot)$ denote the standard normal CDF. In this setting, $x^+ - x^- = \Phi(c) - \Phi(0)$, and weak identification corresponds to small values of c. We use the following values to generate the data: $\beta = 0$, $n = 1000$, $c = 2$ for strong identification and $c = 0.01$ for weak identification, and the values of $\rho = 0.50$ and 0.99.

For each Monte Carlo replication, we generate n observations as described above. Using the bandwidth $h = n^{-1/5-1/100}$ and the uniform kernel, we compute $\hat{\beta}$, $\hat{V}_\beta$, and the confidence intervals $\hat{\beta} \pm z_{1-\alpha/2}\sqrt{\hat{V}_\beta/(nh)}$ for $\alpha = 0.10, 0.05, 0.01$. We use 10,000 replications to compute the average coverage probabilities of those confidence intervals, as well as bias and MSE of the FRD estimator.

The results are reported in Table 1. When identification is strong ($c = 2$), the usual confidence intervals have coverage probabilities very close to the nominal ones. This is regardless of the degree of endogeneity ($\rho = 0.50$ or $\rho = 0.99$). The FRD estimator is more biased when ρ is large.

When identification is weak ($c = 0.01$) and the degree of endogeneity is small ($\rho = 0.50$) the usual confidence intervals include the true value of β with higher than nominal probabilities. As a matter of fact, the obtained coverage probabilities are close to one. This finding is consistent with our analytical results, see the discussion following Theorem 1.

The situation changes when identification is weak ($c = 0.01$) but endogeneity is strong ($\rho = 0.99$). In this case we observe size distortions. For example, the actual coverage probabilities of the 90%, 95% and 99% confidence intervals are approximately 82%, 87% and 95% respectively. We also repeated the experiment with z_i generated as $N(0, 10^2)$. In this case, more substantial size distortions were observed, and the actual coverages of the 90%, 95% and 99% confidence intervals were 76%, 82% and 90% respectively.

When identification is weak, the FRD estimator is also more biased and has large MSE and standard errors. To investigate further the behavior of the FRD estimator, its standard errors, and the usual t-statistic, we estimate their distributions. Figure 1 reports the density of $\hat{\beta}$ estimated by kernel smoothing. It shows that in the case of weak identification, the FRD estimator $\hat{\beta}$ suffers from very heavy tails. Also, when $\rho = 0.99$, the distribution of $\hat{\beta}$ is bimodal.

The distribution of the standard error $\sqrt{\hat{V}_\beta / (nh)}$ also exhibits heavy tails when identification is weak. This explains the large values for the average standard errors under weak identification reported in Table 1. For example, when $c = 0.01$ and $\rho = 0.50$, the median, 75th percentile, and maximum standard error are 3.98, 13.73, and 6.97×10^6 respectively. The standard errors are well-behaved when identification is strong. For example, when $c = 2$ and $\rho = 0.50$, the median, 75th percentile, and maximum standard error are 0.63, 0.77, and 5.59 respectively.

Lastly, Figure 2 shows the densities of the usual T statistic estimated by kernel smoothing. For comparison, we also plot the standard normal density. For $\rho = 0.50$ and strong identification, it is apparent that the standard normal distribution is a very good approximation to the distribution of T. When $\rho = 0.99$, the distribution of T is slightly skewed to the left, but the normal approximation still works reasonable well, because there is no substantial deviation of extreme values of the distribution of T from those of the standard normal distribution.

Figures 2(c) and (d) show that under weak identification, the distribution of T is very different from normal. It is strongly skewed to the left, however, when $\rho = 0.50$ it is also more concentrated around zero. As a result, we do not see size distortions when identification is weak but the degree of endogeneity is small. The picture changes drastically when $\rho = 0.99$. The distribution of T is strongly skewed to the left and no longer concentrated around zero. As a result, we observe size distortions in this case.

We also computed the simulated coverage probabilities of the weak identification robust confidence set $CS_{1-\alpha}$ introduced in (4) in Section 2.3. We found that regardless of the strength of identification c and degree of endogeneity ρ, the simulated coverage probabilities of $CS_{1-\alpha}$ were remarkably close to the nominal coverage probabilities: for the 90%, 95% and 99% confidence sets the simulated coverage probabilities were 90.2%, 95.4% and 99.3% respectively. This supports our claim that the inference based on the null-restricted statistic $\tilde{T}(\beta_0)$ does not suffer from size distortions and is asymptotically valid unlike the testing procedures based on the usual t-statistic.

The form (and the expected length) of the weak identification robust confidence set $CS_{1-\alpha}$ however does depend on the strength of identification. As reported in Table 2, in the case of a strong FRD, the probabilities for the robust confidence sets to be equal to the entire real line are very small. Regardless of the value of ρ, they were below 1% for the 90% and 95% confidence sets, and approximately 2% for the 99% confidence set. The probabilities for the robust confidence to be given by a union of two half lines were similarly small. In the case of a weak FRD, unbounded robust confidence sets were obtained with very high probabilities. Thus, the entire real line was obtained with probabilities 74%, 86% and 97% for the confidence sets with nominal coverage of 90%, 95% and 99% respectively.

Appendix A: Proofs of the theorems

Proof of Theorem 1. For part (a), using (1),

$$\begin{aligned}
\hat{\beta} - \beta &= \frac{\sqrt{nh}\left((\hat{y}^+ - \hat{y}^-) - (y^+ - y^-)\right) - \beta\sqrt{nh}\left((\hat{x}^+ - \hat{x}^-) - (x^+ - x^-)\right)}{\sqrt{nh}\left((\hat{x}^+ - \hat{x}^-) - (x^+ - x^-)\right) + \sqrt{nh}\,(x^+ - x^-)} \\
&= \frac{\sqrt{nh}\left((\hat{y}^+ - \hat{y}^-) - (y^+ - y^-)\right) - \beta\sqrt{nh}\left((\hat{x}^+ - \hat{x}^-) - (x^+ - x^-)\right)}{\sqrt{nh}\left((\hat{x}^+ - \hat{x}^-) - (x^+ - x^-)\right) + \theta} \\
&\to_d \xi_{\beta,\theta},
\end{aligned}$$

where the second equality is by Assumptions 3, and the result in the last line is by Assumption 2 and the Continuous Mapping Theorem.

For part (b), by Assumptions 3,

$$
\begin{aligned}
(nh)^{-1}\hat{V}_{\beta} &= \frac{\hat{\sigma}_{yy}^{+}+\hat{\sigma}_{yy}^{-}+\hat{\beta}^{2}\left(\hat{\sigma}_{xx}^{+}+\hat{\sigma}_{xx}^{-}\right)-2\hat{\beta}\left(\hat{\sigma}_{yx}^{+}+\hat{\sigma}_{yx}^{-}\right)}{nh\left[\left(\hat{x}^{+}-\hat{x}^{-}\right)-\left(x^{+}-x^{-}\right)+\theta/\sqrt{nh}\right]^{2}} \\
&= \frac{\hat{\sigma}_{yy}^{+}+\hat{\sigma}_{yy}^{-}+\hat{\beta}^{2}\left(\hat{\sigma}_{xx}^{+}+\hat{\sigma}_{xx}^{-}\right)-2\hat{\beta}\left(\hat{\sigma}_{yx}^{+}+\hat{\sigma}_{yx}^{-}\right)}{\left[\sqrt{nh}\left(\left(\hat{x}^{+}-\hat{x}^{-}\right)-\left(x^{+}-x^{-}\right)\right)+\theta\right]^{2}} \\
&\to_{d} \frac{\sigma^{2}\left(\beta+\xi_{\beta,\theta}\right)}{\left(X^{+}-X^{-}+\theta\right)^{2}}.
\end{aligned}
$$

For part (c), by imposing $\beta=\beta_0$, collecting the results from (a) and (b), and since convergence in (a) and (b) is joint, we obtain

$$
T\left(\beta_0\right)\to_{d}\operatorname{sgn}\left(X^{+}-X^{-}\right)\frac{\left(Y^{+}-Y^{-}-\beta\left(X^{+}-X^{-}\right)\right)}{\sigma\left(\beta\right)}\frac{\sigma\left(\beta\right)}{\sigma\left(\beta+\xi_{\beta,\theta}\right)}.
$$

where, $\left(Y^{+}-Y^{-}-\beta\left(X^{+}-X^{-}\right)\right)/\sigma\left(\beta\right)\sim N\left(0,1\right)$. □

Proof of Theorem 2. The absolute value of the null-restricted t-statistic $\left|\tilde{T}\left(\beta_0\right)\right| = \left|\hat{\beta}-\beta\right|/\sqrt{\tilde{V}_{\beta}\left(\beta_0\right)/\left(nh\right)}$ can be written as follows:

$$
\begin{aligned}
&\frac{\sqrt{nh}\left|\hat{y}^{+}-\hat{y}^{-}-\beta_0\left(\hat{x}^{+}-\hat{x}^{-}\right)\right|}{\hat{\sigma}\left(\beta_0\right)} \\
&= \frac{\sqrt{nh}\left|\left(\hat{y}^{+}-\hat{y}^{-}\right)-\left(y^{+}-y^{-}\right)-\beta_0\left(\left(\hat{x}^{+}-\hat{x}^{-}\right)-\left(x^{+}-x^{-}\right)\right)+\left(\beta-\beta_0\right)\theta/\sqrt{nh}\right|}{\hat{\sigma}\left(\beta_0\right)} \\
&\to_{d}\left|\mathcal{Z}+\frac{\theta\left(\beta-\beta_0\right)}{\sigma\left(\beta_0\right)}\right|.
\end{aligned}
$$

□

Appendix B

Here we show that the robust confidence set $CS_{1-\alpha}$ defined in (4) cannot be empty. From (5) it follows that for $CS_{1-\alpha}$ to be empty, the following two conditions must be satisfied:

$$nh(\hat{x}^+ - \hat{x}^-)^2(\hat{\beta}^2\hat{\sigma}_{xx} - 2\hat{\beta}\hat{\sigma}_{yx} + \hat{\sigma}_{yy}) - z^2_{1-\alpha/2}\left(\hat{\sigma}_{xx}\hat{\sigma}_{yy} - \hat{\sigma}^2_{yx}\right) < 0, \quad (7)$$

$$nh(\hat{x}^+ - \hat{x}^-)^2 - z^2_{1-\alpha/2}\hat{\sigma}_{xx} > 0. \quad (8)$$

Suppose that the first inequality holds. Since the variance-covariance matrix composed of $\hat{\sigma}_{xx}$, $\hat{\sigma}_{yy}$, and $\hat{\sigma}_{yy}$ is positive definite, it follows that $\hat{\beta}^2\hat{\sigma}_{xx} - 2\hat{\beta}\hat{\sigma}_{yx} + \hat{\sigma}_{yy} > 0$ and $\hat{\sigma}_{xx}\hat{\sigma}_{yy} - \hat{\sigma}^2_{yx} > 0$. The inequality in (7) then can be re-written as

$$\begin{aligned} nh(\hat{x}^+ - \hat{x}^-)^2 &< z^2_{1-\alpha/2}\hat{\sigma}_{xx}\frac{\hat{\sigma}_{yy} - \hat{\sigma}^2_{yx}/\hat{\sigma}_{xx}}{\hat{\beta}^2\hat{\sigma}_{xx} - 2\hat{\beta}\hat{\sigma}_{yx} + \hat{\sigma}_{yy}} \\ &< z^2_{1-\alpha/2}\hat{\sigma}_{xx}\left(1 - \frac{(\hat{\sigma}_{yx}/\sqrt{\hat{\sigma}_{xx}} - \hat{\beta}\sqrt{\hat{\sigma}_{xx}})^2}{\hat{\beta}^2\hat{\sigma}_{xx} - 2\hat{\beta}\hat{\sigma}_{yx} + \hat{\sigma}_{yy}}\right) \\ &< z^2_{1-\alpha/2}\hat{\sigma}_{xx}. \end{aligned}$$

It follows that the two inequalities (7)-(8) cannot be true together.

References

Anderson, T. W., and H. Rubin (1949): "Estimation of the parameters of a single equation in a complete system of stochastic equations," *The Annals of Mathematical Statistics*, 20(1), 46–63.

Andrews, D. W. K., and J. H. Stock (2007): "Inference with Weak Instruments," in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, ed. by R. Blundell, W. K. Newey, and T. Persson, vol. III. Cambridge University Press, Cambridge, UK.

Dufour, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1387.

Hahn, J., P. Todd, and W. Van der Klaauw (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69(1), 201–209.

Imbens, G. W., and T. Lemieux (2008): "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142(2), 615–635.

Otsu, T., and K.-L. Xu (2010): "Empirical Likelihood for Regression Discontinuity Design," Working Paper.

Staiger, D., and J. H. Stock (1997): "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65(3), 557–586.

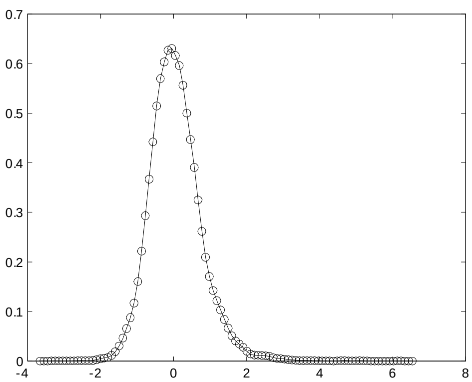
Table 1: Simulated coverage probabilities of the confidence intervals constructed as estimate $\pm$ constant $\times$ standard error, and bias, root MSE, and average standard error of the FRD estimator

identification	endogeneity	nominal coverage	simulated coverage	bias	root MSE	std.err.
strong	$\rho = 0.50$	0.90	0.9307	0.0557	0.7100	0.6987
		0.95	0.9710			
		0.99	0.9951			
strong	$\rho = 0.99$	0.90	0.9264	0.1114	0.8021	0.7360
		0.95	0.9563			
		0.99	0.9858			
weak	$\rho = 0.50$	0.90	0.9803	-1.0765	72.2496	2.1235×10^3
		0.95	0.9927			
		0.99	0.9995			
weak	$\rho = 0.99$	0.90	0.8219	-0.1356	133.3221	1.3184×10^4
		0.95	0.8749			
		0.99	0.9459			

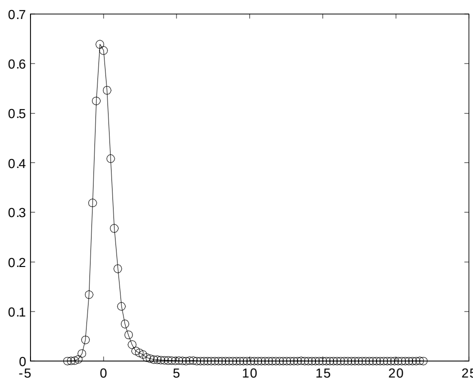
Table 2: Simulated probabilities for the weak identification robust confidence set $CS_{1-\alpha}$ to be the entire real line or a union of two disconnected half-lines

identification	endogeneity	nominal coverage	entire real line	two half-lines
strong	$\rho = 0.50$	0.90	0.0017	0.0033
		0.95	0.0046	0.0080
		0.99	0.0213	0.0271
strong	$\rho = 0.99$	0.90	0	0.0061
		0.95	0.0004	0.0122
		0.99	0.0025	0.0455
weak	$\rho = 0.50$	0.90	0.7441	0.1579
		0.95	0.8581	0.0953
		0.99	0.9719	0.0211
weak	$\rho = 0.99$	0.90	0.7444	0.1578
		0.95	0.8599	0.0963
		0.99	0.9707	0.0227

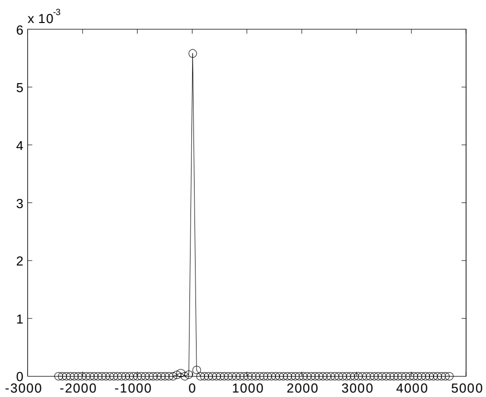
Figure 1: Kernel estimated density of the FRD estimator under strong and weak identification for different values of the degrees of endogeneity ρ



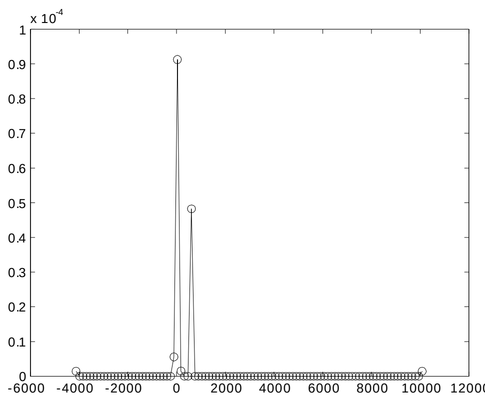
(a) Strong identification, $\rho = 0.50$



(b) Strong identification, $\rho = 0.99$

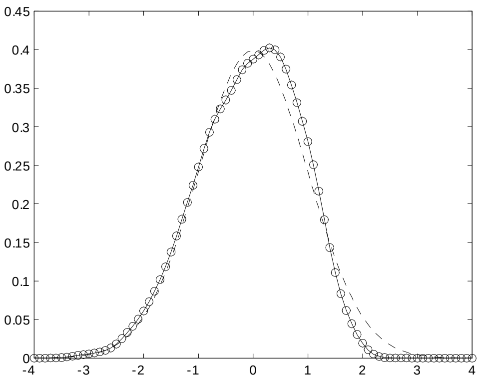


(c) Weak identification, $\rho = 0.50$

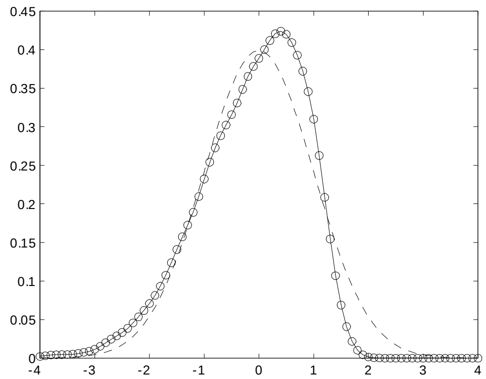


(d) Weak identification, $\rho = 0.99$

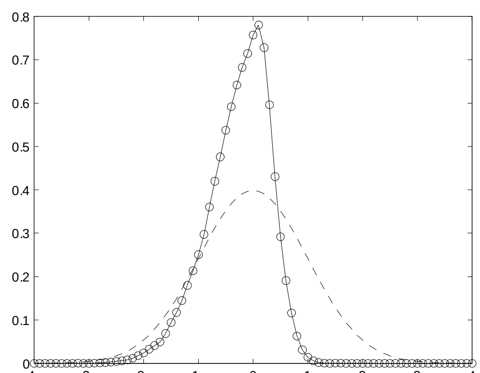
Figure 2: Kernel estimated density of the T statistic (solid line) under strong and weak identification for different values of the degrees of endogeneity ρ against the standard normal PDF (dashed line)



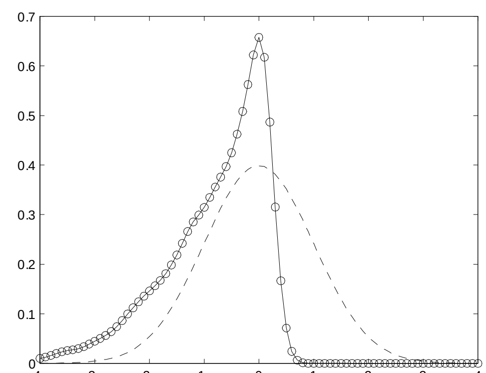
(a) Strong identification, $\rho = 0.50$



(b) Strong identification, $\rho = 0.99$



(c) Weak identification, $\rho = 0.50$



(d) Weak identification, $\rho = 0.99$